

Towards artificial general intelligence via a multimodal foundation model

Year: 2021 | Citations: 272 | Authors: Nanyi Fei, Zhiwu Lu, Yizhao Gao, Guoxing Yang, Yuqi Huo

Abstract

The fundamental goal of artificial intelligence (AI) is to mimic the core cognitive activities of human. Despite tremendous success in the AI research, most of existing methods have only single-cognitive ability. To overcome this limitation and take a solid step towards artificial general intelligence (AGI), we develop a foundation model pre-trained with huge multimodal data, which can be quickly adapted for various downstream cognitive tasks. To achieve this goal, we propose to pre-train our foundation model by self-supervised learning with weak semantic correlation data crawled from the Internet and show that promising results can be obtained on a wide range of downstream tasks. Particularly, with the developed model-interpretability tools, we demonstrate that strong imagination ability is now possessed by our foundation model. We believe that our work makes a transformative stride towards AGI, from our common practice of “weak or narrow AI” to that of “strong or generalized AI”. Artificial intelligence approaches inspired by human cognitive function have usually single learned ability. The authors propose a multimodal foundation model that demonstrates the cross-domain learning and adaptation for broad range of downstream cognitive tasks.